The Need for Resilience Research in Coupled Big Compute and Big Data Scientific Applications

January 20, 2014

Franck Cappello and Tom Peterka, Argonne National Laboratory

Projections and reports about exascale failure modes conclude that we need to protect numerical simulations and data analytics from an increasing risk of hardware and software failures and silent data corruptions (SDC). At this scale, hardware and software failures could be as frequent as several per hour. According to [1], the semiconductor industry will have increased difficulty presenting software with an efficient dependable hardware layer when feature size will become lower than 10nm (11nm is projected in 2015-2017 according to Intel and NVIDIA). For coupled computation and data analytics at extreme scale, the challenge is to produce correct results in the presence of potentially unreliable hardware and software.

After approximately one dozen workshops and reports on exascale resilience, the need for resilience of parallel computations at extreme scale (big compute) is widely accepted. Roadmaps give priority to improving checkpoint restart, developing new programming models and runtimes for resilience, and focusing on detection to limit SDC as much as possible. These roadmaps were primarily concerned with ensuring that large-scale simulations complete and produce correct results. In other contexts such as clouds and grids (big data), many research results concern resilience of workflows, including computation, data storage, and data analytics. However, we are not aware of such roadmaps or research efforts specifically concerning resilience of coupled simulation and data analytics (big compute + big data) in the context of scientific applications on extreme-scale platforms such as the future exascale systems.

Coupled big compute and big data scientific applications have characteristics that distinguish them from single-component scientific simulation, coupled numerical simulation, and workflows running on grids and clouds. Coupled big compute and big data scientific workflows connect producer and consumer parallel components in complex data flow graphs. Figure 1 shows an example workflow coupling a cosmology simulation with a small subset of associated data analytics. The ovals represent analysis and visualization programs that convert data to other forms (shown in squares). For example, the simulation code, *HACC* [1], produces raw particles that can be viewed directly with *ParaView*, a production visualization tool. Or, particles can be converted to a mesh tessellation through the *tess* [2] parallel library, which can be visualized or further resampled onto a regular grid with the *dense* parallel tool. Other serial utilities can operate offline on the tessellation and density fields.

These executions are much less regular than standalone scientific simulations: components may use different computing and communication models; communication occurs at multiple levels (within components and between components); orchestrators organize the execution of the workflow, and data flow may be scheduled to increase performance. Analysis workflows also differ from coupled simulations (for example multiphysics codes) because (i) a stream of data is produced by the scientific simulation, and multiple stages transform the data in the stream; (ii) components are of different nature: numerical simulations, data transposition, data reductions, visualizations; (iii) ultimately the data is represented in ways that promote scientific understanding. The resilience expectations depend on these factors. For example, if the workflow

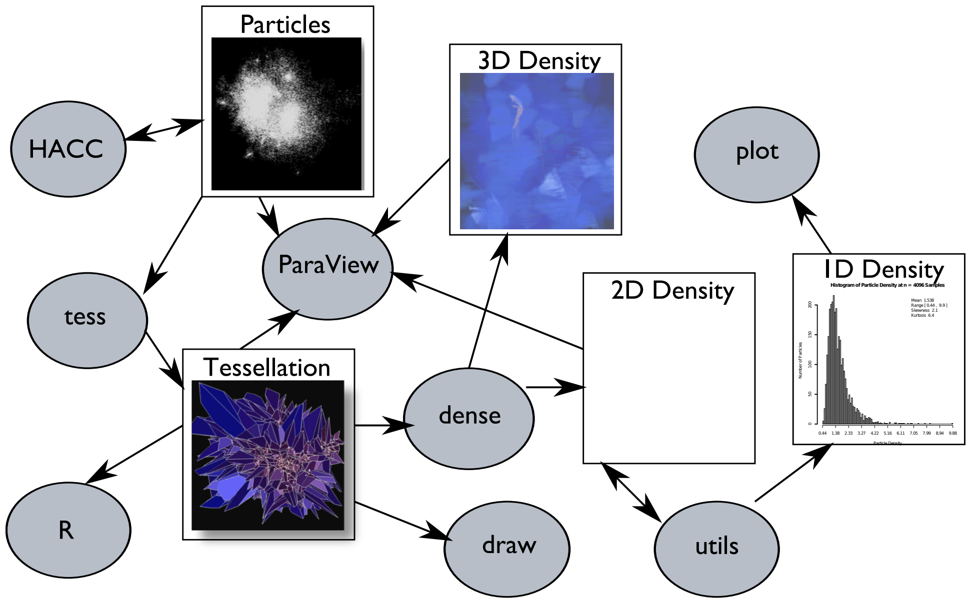

Figure 1: A workflow coupling a cosmology simulation with its associated data flow for the conversion of particle data to unstructured and regular grid analysis products. The ovals represent programs, some parallel and others serial, both tightly and loosely coupled. Raw and derived data products are denoted by squares.

produces an animation, it may be acceptable to lose several video frames because of a transient component failure; the large amount of compression and interframe coherence in the movie may entirely hide the missing data.

Other differences exist between workflows run on grid/cloud platforms and coupled big compute + big data analytics workflows on extreme-scale platforms. Compared to grids/clouds, extreme-scale platforms are less heterogeneous; resources are supposed to be more reliable at a comparable scale; there are fewer security concerns (data are accessed within a single administration domain); there is less orchestration concern because resources are easier to reserve and allocate, and communication performance is orders of magnitude superior.

Other major differentiators between extreme-scale platforms and grid/cloud environments are the type and scale of workflows. In grid and cloud platforms, workflows execute over thousands of cores, while at extreme scale, we expect to execute workflows over millions of cores. Grid and cloud platforms typically are loosely coupled, where data is stored to disk after each transformation, avoiding direct communication between components for performance and reliability reasons. Extreme-scale platforms will run tightly coupled workflows and will use high-performance communication support (including, in memory communication) to perform direct communication between computational and data analytics components.

The loose coupling of grid and cloud workflows affords failure containment between workflow components and the possibility to restart failed components individually (under some assumptions) from intermediary stored data. It is unlikely that workflow for coupled simulation and data analytics on extreme scale platform will have such fault containment properties. On the contrary, we believe that the performance objectives will orient workflow executions on these platforms toward tightly coupled executions where high performance communications are used to exchange information between components running concurrently and where dependencies between components can lead to cascading effects when failures happen.

Another important aspect of extreme scale workflows is the probability and importance of silent data corruption (SDC). To our knowledge, there is no published research concerning SDC even in grid/cloud workflows, and we expect the probability of SDC to be higher at extreme-scale, justifying research on

SDC detection. The consequences of SDC in analysis warrants further research as well: depending on the "importance" of a data product, the combination of resolution and location in the workflow make some data products more sensitive to SDC than others. For example, referring back to Figure 1, the raw particles and tessellation are high resolution data, not visual or statistical summaries, and several hops separate them from sinks in the workflow. We hypothesize that therefore they are more important to protect from SDC, but in general the effect of SDC on various data products in such data flow networks is poorly understood today.

In summary, these differences with single-component extreme scale simulations, with coupled scientific simulations, with grid and cloud workflows, and the complex effects of SDC on analysis products imply that fault tolerance for coupled big compute and big data analytics at extreme scale should be considered as a novel research topic.

References

[1] Salman Habib, Vitaly Morozov, Hal Finkel, Adrian Pope, Katrin Heitmann, Kalyan Kumaran, Tom Peterka, Joe Insley, D. Daniel, Pat Fasel, Nick Frontiere, and Z. Lukic. The Universe at Extreme Scale: Multi-Petaflop Sky Simulation on the BG/Q. In *Proceedings of SC12*, Salt Lake City, UT, 2012.

[2] Tom Peterka, Juliana Kwan, Adrian Pope, Hal Finkel, Katrin Heitmann, Salman Habib, Jingyuan Wang, and George Zagaris. Meshing the Universe: Integrating Analysis in Cosmological Simulations. In *Proceedings of the SC12 Ultrascale Visualization Workshop*, Salt Lake City, UT, 2012.